

A Thesis, presented as part of the requirements for the Award of a Research Masters  
Degree in Economics from the NOVA – School of Business and Economics.

The reasons behind the progression in PISA scores: An  
education production function approach using  
semi-parametric techniques

Pedro Freitas, Student number 38

A project carried out on the Research Masters in Economics, under the supervision of:

Prof. Ana Balcão Reis

# The reasons behind the progression in PISA scores: An education production function approach using semi-parametric techniques \*

Pedro Freitas <sup>†</sup>

*Nova School of Business and Economics, Universidade Nova de Lisboa*

April 30, 2015

## Abstract

On December 2013, and for the fifth time since 2000, OECD published the results of the latest PISA survey, providing a view on how the students' performance has progressed during the last 12 years. Using PISA data we follow an education production function, which states that variables related to students, their family and the school explain the output, measured as the individual student achievement. Exploring the concept of efficiency we measure the ability that each student has to transform the given inputs into higher academic outcomes. Such analysis was performed through the estimation of an efficient frontier, derived by non-parametric techniques, namely Data Envelopment Analysis (DEA). Using this methodology we establish two vectors of analysis. The first one intends to disentangle the reasons behind the evolution in PISA scores across the years, concluding that the variation in inputs is on the core of the reasons to explain the evolution in PISA results. The second aims to evaluate what are the sources of student's efficiency. On this topic we particularly explore the role of the school inputs, concluding that students with a more favourable socio-economic background are more indifferent to variables such as class size and school size.

*JEL classification:* I21,C61,C67.

*Keywords:* Education production function; Efficiency; Data envelopment analysis; PISA.

---

\*I am greatly indebted to my supervisor, Ana Balcão Reis, for the unexcelled guidance and encouragement and whose advice enormously improved this work. I acknowledge the computational help given by Luís Catela Nunes. I thank as well José Manuel Cordero-Ferrera for the help with Matlab coding. I appreciate the comments from the participants in the research group of Nova SBE, in the XXIII Meeting of the Economics of Education Association, Valencia 2014 and in the Third Lisbon Research Workshop on Economics, Statistics and Econometrics of Education. A reference as well to the support and comments from my Research Master and Phd colleagues namely, Marta Lopes, João Santos, Marta Rocha, Catarina Ângelo, Sofia Oliveira and Maria Ana Matias. Finally a very special acknowledgment to João Vaz, whose friendship, support and wisdom were so important during all this time and who shows in every moment the truth in Beckett's sentence: "Friendship, according to Proust, is the negation of that irremediable solitude to which every human being is condemned."

<sup>†</sup>pedro.freitas@novasbe.pt

# 1 Introduction

PISA is a set of three tests in the fields of reading, mathematics and sciences targeting students between 15 and 16 years old and organized every three years by OECD since 2000. Taking advantage of the availability of data covering a period of 12 years we intend to address two issues; identify the drivers behind the evolution in PISA scores and assess the reasons that justify the students' relative level of efficiency. In our specific case we have used data from Portugal, which has been a participating country in PISA since 2000, showing a consistent, but not constant evolution during this time period.<sup>1</sup>

Regarding the motives behind the evolution in PISA scores, previous work tried to identify them, particularly Pereira (2011) for Portugal and Osorio *et al.* for Indonesia (2011). Both papers followed an Oaxaca-Blinder decomposition, which consists in explaining the change in PISA scores through linear regressions and observe if this is attributed to the explanatory variables or to the evolution in the regression coefficients.<sup>2</sup>

Instead of following this decomposition approach we opted for a frontier analysis, which adds an efficiency perspective to the education production function, deriving relative measures across the students. Under such analysis, we propose that students combine a set of variables related to them, to their families and to their school (e.g. the inputs) in order to obtain a certain level of academic achievement (e.g. the output). The ones who use these variables in the most efficient way, define what is named as the efficient frontier, and all the remaining students are considered as inefficient. Thus this methodology enables us to define a standard for comparing different students conditioning on their specific characteristics and their academic results.

To construct the efficient frontier two paths can be followed: 1. A parametric estimation, following a stochastic frontier analysis; 2. A non-parametric approach, namely through the use of Data Envelopment Analysis (DEA). Examples of the first approach can be found in the literature in works such as Pereira and Moreira (2011), which studies the efficiency of the Portuguese secondary schools, Perelman and Santín (2005), on the efficiency of the Spanish students using PISA data or Ryan (2013) who uses a stochastic frontier analysis to identify the reasons for the decline in Australian PISA scores.

Given the puzzling results often obtained from standard parametric estimations, the use of non-parametric approaches has gained attention in the literature, as referred by Worthington (2001), who points out two main reasons for the attractiveness of these methodologies. By one side it is mentioned the freedom in the relation between inputs

---

<sup>1</sup>A graphical summary of the evolution of the Portuguese performance in PISA can be seen in Appendix A<sub>1</sub>. Appendix A<sub>2</sub> presents a brief view on the structure of the Portuguese education system.

<sup>2</sup>Pereira (2011) concluded that from 2003 to 2006 the evolution in the variables level was the dominant effect, while from 2006 to 2009 both effects contributed positively to the evolution of the PISA results.

and outputs, particularly in education contexts where some of the usual axioms of productivity may breakdown (e.g. cost minimization). It is referred as well that many times in standard statistical approaches the desired set of data is not available, which makes approaches that impose few restrictions on data, such as DEA, more attractive.

Thus, we propose a DEA model previously used in economics of education literature, but contrasting to previous applications of this methodology we have enlarged the analysis in three particular ways: 1. We have followed a regression based election of inputs, in order to reduce the level of *ad-hoc* choice of the inputs to be included in the frontier estimation, 2. We have used data from several periods and compared the results across time; 3. Since we aim to center the analysis at the student level, we are able to derive an efficiency score for each one of the students who participated in PISA since 2003, allowing us to identify the position of each student in relation to the efficient frontier.

More than simply stating what are the reasons behind the evolution in PISA scores we aim to use the frontier approach to explain the sources of the students' efficiency. To this propose we have divided the analysis into two main parts: the first explains what are the socio-economic environmental factors behind the relative efficiency of each student, namely the wealth of family or the parents' schooling level. Secondly we evaluate in which way the allocation of school resources contribute to the students' efficiency, namely who are the students that have been able to take more advantage of the school resources available. In the context of school inputs we also purpose an approach, based on Data Envelopment Analysis techniques, to assess how a different allocation of school resources, in particular school size and class size, may change or not the students' achievement.

Section 2 provides a theoretical framework to the question studied; section 3 presents the PISA data used; sections 4 describes the methodology and the results on the reasons behind the evolution in PISA scores, section 5 presents the factors that determine the students' efficiency and finally section 6 concludes.

## 2 Theoretical background

Todd and Wolpin (2003) present an exhaustive theoretical exposition on the education production function, particularly on how to model the relation between socio-economic variables and the individual student achievement, following previous work by Hanusheck (1986, 2003). As it is referred in this literature, ideally when assessing the student's performance we would dispose of all the past history of the variables used, whether they relate to the school or to the student's family. However in many real data applications, we do not have access to such a wide set of information. Particularly, that is the case of the PISA survey which observes a given student just once (when he is around 15 years

old) leading us to follow what Todd and Wolpin call a contemporaneous model, where the variables used to explain the output (the student's academic performance) should be able to capture the past history of inputs. Considering a certain student,  $i$ , which lives in household,  $j$ , and studies in school,  $s$ , we assume:

$$T_{ijs} = T(St_{ijs}; F_{ijs}; S_{ijs}) + \epsilon_{ijs} \quad (1)$$

Where  $T_{ijs}$  is the student's achievement, measured by the score in a particular test, and it is a function of: 1. Student's characteristics,  $St_{ijs}$ ; 2. Family inputs,  $F_{ijs}$ ; 3. School inputs,  $S_{ijs}$ ; 4. An additive error, which includes all the unobserved characteristics of the individual,  $\epsilon_{ijs}$ .

Building on this specification we propose a similar framework, to which we add a productivity factor,  $A_{ijs}$ , standing for the ability students have to use the set of inputs available to generate their output:

$$T_{ijs} = T(A_{ijs}, St_{ijs}; F_{ijs}; S_{ijs}) \quad (2)$$

We combine this concept of education production function with the notion of efficiency. Koopmans (1959) defined technical efficiency as the capacity to maximize the output conditional on the level of the inputs. The agents who can achieve such maximization are the ones who define the efficient frontier, and as Farrell (1957) referred inefficiency is determined as deviations from this same frontier. Thus, the set of efficient points, or in our case efficient students, are the ones who are on the frontier and the inefficient are the ones who are below the frontier.

Using this approach we have derived the education production frontier, and use it to study the three possible reasons to explain the evolution in PISA results: 1. An overall improvement in efficiency, meaning that more students are closer to the frontier; 2. A shift in the education frontier (e.g. the students' ability to transform inputs in outputs); 3. The inputs related to the student, his family and the school increased.<sup>3</sup>

---

<sup>3</sup>In the proposed education production function we assume positive marginal returns.

These three hypotheses are tested, concluding on which ones seem more plausible to explain the Portuguese case.<sup>4</sup> Among the three reasons presented, the first two correspond to a shock in the term,  $(A_{ijs})$ , considering the specification in (2).<sup>5</sup>

### 3 Data

We have used the PISA dataset, which has the virtue of having a rich set of variables gathered through several questionnaires designed by OECD. There are 5 PISA surveys available from 2000 to 2012, although the initial year was excluded from our analysis due to the difference between the number of individuals who performed the reading and mathematics test, remaining so with 4 PISA samples (2003, 2006, 2009 and 2012).

The survey intends to be a representative sample of the 15 years old student population at the time it is performed. To preserve this representativeness criteria, PISA dataset includes weights for each observation which, as recommended by OECD, are taken in consideration in all our steps to support the results.<sup>6</sup>

In order to meet the theoretical formulation presented in equations (1) and (2), the variables chosen for the analysis were grouped in factors attributed to the student, the student's family and the school. Since we are working with data from several years, the choice of variables was subject to the existence and comparability across several PISA surveys. For example two variables that could be taken as school inputs, the index Respres (which stands for the level of autonomy in resources management) or the percentage of repeaters at the school had to be dropped from the analysis, the first because its scale of measurement changed from 2003 to 2006 and the latter due the fact that such question was not included in the schools questionnaires in 2012.

When treating the data we took advantage of the indexes released by PISA, which aggregate several answers given by students and schools in the questionnaires.<sup>7</sup> These indexes reveal to be important to define proxies, for example for students' family economic situation, which is measured by the Index of Home Possessions (Homepos). PISA also release the family Index of Cultural Possessions (Cultpos) and Educational Resources at home

---

<sup>4</sup>A graphical illustration of these possibilities can be seen in Appendix A3.

<sup>5</sup>Translating these three hypothesis into the case of an Oaxaca-Blinder decomposition, the first two cases correspond to a change in the coefficient and the third case to a change in the variables level. Through the proposed approach it is possible to derive a more refined analysis that trough an Oaxaca-Blinder methodology, since the evolution in coefficients is divided into two possible sources.

<sup>6</sup>The methodology followed by OECD to construct and weight the sample can be found in PISA Data Analysis Manual and PISA Technical Reports.

<sup>7</sup>Todd and Wolpin (2003) advice against the use of proxies. However its use reveals to be necessary in order to control for the student's social and economic background.

(Hedres). We have used these three indexes all together to measure the proportional investment of the student's family in cultural possessions and educational resources given by the variables, *Cultposinv* and *Hedresinv* respectively:

$$Cultposinv = \frac{Cultpos}{Homepos} \quad (3)$$

$$Hedresinv = \frac{Hedres}{Homepos} \quad (4)$$

Attention was also given to existence of missing data. The percentage of missing data by variable is relatively low, but when combining several variables for empirical estimation, the sample can be significantly reduced (e.g. considering the variables chosen to the analysis, the samples are reduced for values between 50% in 2012 and 85% in 2009).<sup>8</sup> In order to address this same problem, we have opted for the implementation of data imputation techniques, which can be divided in single and multi imputation. This last one was excluded since we need a single imputation for each missing observation, to apply the DEA method followed after. It was run an Expected Maximization algorithm<sup>9</sup> to fill part of the missing data, making that the final imputed samples include at least 90% of the observations of the original sample.<sup>10</sup>

The variables which resulted from all the data treatment and the respective descriptive analysis can be observed in table below;

---

<sup>8</sup>Sample size without missing answers: 3462 (2003); 4014 (2006), 5535 (2009); 2831 (2012).

<sup>9</sup>The explanation of this same algorithm can be found in appendix A4.

<sup>10</sup>Sample size after imputation: 4252 (2003); 4805(2006); 5984(2009);5178 (2012).

Table 3 .1<sup>11</sup>

(weighted sample 2003-2012)	PISA variable	Interpretation	Mean	Sd	Min	Max
<b>Student's variables, <math>St_{ijs}</math></b>	8th grade (dummy=1)	-	0.09	0.28	0	1
	9th grade (dummy=1)	-	0.25	0.43	0	1
	10th grade (dummy=1)	-	0.58	0.49	0	1
	11th grade (dummy=1)	-	0.004	0.061	0	1
	Academic Track (dummy=1)	Students enrolled in the standard academic track	0.87	0.34	0	1
	Female (dummy=1)	-	0.51	0.49	0	1
	Age	-	15.78	0.29	15	16.25
<b>Family variables, <math>F_{ijs}</math></b>	FISCED	Father's position in the international standard classification of education, scale 0-6	2.52	2.03	0	6
	MISCED	Mother's position in the international standard classification of education, scale 0-6	2.49	2.04	0	6
	Parents pt1 (dummy=1)	Students whose one of parents were born in Portugal	0.12	0.33	0	1
	Parents pt2 (dummy=1)	Students whose both parents were born in Portugal	0.83	0.37	0	1
	CULTPOSinv	Index of investment in cultural possessions at home	0.27	0.12	0.06	0.84
	HEDRESinv	Index of investment in educational resources	0.72	0.11	0.11	1.84
	Homepos	Index for level of home possession at home	7.2	0.98	0.58	11.21
<b>School variables, <math>S_{ijs}</math></b>	Non gov(%)	Percentage of non-governmental financing of the school	16.75	22.5	0	100
	Schools in small towns (dummy=1)	-	0.31	0.46	0	1
	Schools in towns (dummy=1)	-	0.4	0.49	0	1
	Schools in cities (dummy=1)	-	0.16	0.36	0	1
	Schools in large cities (dummy=1)	-	0.05	0.22	0	1
	Percentage of girls in schools	-	50.46	5.24	0.6	100
	School size	-	978.31	474.14	73	2750
	Students/teacher ratio	-	9.33	4.08	0.885	68
	Scmatedu	Index for level of quality of school resources	2.89	0.87	0.26	5.2
	Class size <sup>12</sup>	-	22.52	4.66	1	53
	Proportion of certified teachers	-	89.01	18.46	0.6	100
	Tschort	Index for teacher shortage	1.26	0.6	0.79	5.59
	N	Number of students	20219	-	-	-
	Nw	Weighted Number of students	353,443	-	-	-

<sup>11</sup>Mean and standard deviations calculated based on the imputed sample.<sup>12</sup>For 2006 it corresponds to the average class size per school; and for the other years to the number of students reported by the own students in their classes.



Regarding the results obtained by the students in the PISA tests from the three subjects evaluated we focused our attention in the reading and mathematics ones<sup>13</sup>, denoting a negative evolution in the results from 2003 to 2006 and then a continuous progression between 2006 and 2012, that is more pronounced between 2006 and 2009:

Table 3.2<sup>14</sup>

(weighted sample)	2003	$\Delta$	2006	$\Delta$	2009	$\Delta$	2012
<b>Reading</b>	481.05	-4.76	476.29	15.63	491.92	2.29	494.21
<b>Mathematics</b>	469.8	-0.44	469.36	20.03	489.39	4.25	493.64

## 4 Methodology and Results

The construction of the efficient frontier is performed through Data Envelopment Analysis, whose non-parametric procedure has its roots in the paper by Charnes, Cooper and Rhodes (1978). The intensive use of DEA approaches to measure efficiency in education is found in works such as the ones by Kirjavainen and Loikkanen (1998) on the efficiency of the Finnish secondary schools, Portela (2011, 2012) on the efficiency of Portuguese secondary schools or Afonso and St Aubyn (2005) on cross country efficiency of secondary education.

The wide majority of the literature applying DEA to education focus on the measurement of school efficiency, taking each school as a Decision Making Unit (DMU) (e.g. the entity responsible for converting inputs into outputs). We want to measure how each individual takes advantage of the available inputs, centring our analysis at the student level. This approach has several advantages, since it allows for an efficiency measurement for each individual student, constructing a DEA model with a significantly larger sample than the ones normally used. This path of study is much less usual in the literature, with few exceptions such as Waldo (2005), who applies DEA techniques to student individual data from the Swedish upper secondary schools.

For efficiency calculation it was pursued a DEA output oriented model with variable returns to scale based on the following linear programming problem:

<sup>13</sup>We have decided not to focus in Sciences since Reading and Mathematics are the two fields which are subject to national exams at the end of 9th grade, when students are around 14/15 years old.

<sup>14</sup>These scores do not necessarily match the ones officially reported by PISA, since even after the imputation not all the PISA sample was used.

$$\begin{aligned}
& \max \theta_0 \\
& s.to : \\
& \theta_0 y_0 \leq \sum_{i=1}^n y_i \lambda_i \\
& x_0 \geq \sum_{i=1}^n x_i \lambda_i \\
& \sum_{i=1}^n \lambda_i = 1 \\
& \lambda_i \geq 0
\end{aligned} \tag{5}$$

The problem presented above is solved for  $i+1$  DMU's, having each one the outputs  $y_0$  and the inputs  $x_0$ . This formulation stands for an output oriented model, where efficiency is read as the maximum amount of output that each DMU is able to generate with the inputs given. An efficient DMU is the one who is on the frontier which happens when  $\theta_0 = 1$ , being it the minimum possible value of  $\theta_0$ . All the efficiency scores higher than one correspond to inefficient units, and the higher the value of  $\theta_0$ , the less efficient is the respective DMU.<sup>15</sup> The attractiveness of the DEA approach is that each unit is seen as a linear combination of the most efficient units closest to it, and the weight given to each one of this efficient units is given by the term  $\lambda_i$ . This way the performance of each student is evaluated in relation to his peers. The assumption that the sum of all  $\lambda_i$  equals one imposes variable returns to scale to our problem, as widely used in the literature.<sup>16</sup> Translating the general DEA framework to the specific case of our data, it is assumed a multiple output for each DMU, given by the score achieved in the reading and mathematics test.<sup>17</sup> Regarding the inputs choice we have chosen a methodology that decreased the level of discretion in its election. In DEA original formulations no particular methods were advanced to determine what inputs may or may not be relevant in the construction of the production frontier.<sup>18</sup> In line with what was previously presented in Mancebón-Torrubia *et al.* (2010) we have adopted a regression based methodology for the inputs choice, where estimation analysis were performed using PISA data and following the general theoretical formulation presented in equation (1). OLS regressions were

---

<sup>15</sup>For example if the value of  $\theta_0$  for a certain DMU equals 1.2, it means that with the same inputs, that student could achieve a 20% higher output.

<sup>16</sup>A graphical representation of the frontier possibility set can be seen in Appendix 5.

<sup>17</sup>These scores correspond to the average of the 5 plausible values released by PISA in both fields, Reading and Mathematics.

<sup>18</sup>Later literature on the topic of inputs choice in DEA can be found in works such as Nataraja and Johnson (2011).

computed (as presented in equation 6) taking in consideration the 5 Plausible values and the population weights provided by PISA. The variance of the estimators was estimated by the Balanced Repeated Replication using the 80 population weights provided in the datasets and a Fay's adjustment of 0.5, as recommended by PISA.<sup>19</sup> The estimation was run separately for the reading and mathematics scores considering different cases: 1. Just the students without missing answers (2003-2012); 2. The imputed databases without control for year adjustments (2003-2012); 3. The imputed databases controlled for year dummies (2003-2012).

$$T_{ijs} = \beta_0 + \beta_1 St_{ijs} + \beta_2 F_{ijs} + \beta_3 S_{ijs} + \epsilon_{ijs} \quad (6)$$

Additionally, and using the imputed sample controlled for year effects, it was performed an estimation following an Hierarchical Linear Model (HLM). The use of multilevel models has been growing in the economics of education literature, in works such as Mancebón-Torrubia *et al.* (2010), Bowers and Urick (2011) or Agasisti and Cordero-Ferrera (2013). The PISA own publication already mentioned the possibility of using the PISA sample as a way to perform multilevel analysis.<sup>20</sup> Assessing student performance is one of the examples that best fits the intention of an HLM, which controls for the fact that the observations are organized in several hierarchical structures (e.g each student is nested in a certain class, which is nested in a school, which belongs to a particular region). In our particular case we follow a two level model, being the first level the one correspondent to the student and the second one the one correspondent to the school. All slopes are fixed and only the intercept was randomized at the second level.<sup>21</sup>

#### **Level 1-Model:**

$$T_{ijs} = \beta_{0s} + \beta_1 St_{ijs} + \beta_2 F_{ijs} + \epsilon_{ijs} \quad (7)$$

---

<sup>19</sup>Further details on this methodology can be found in pages 79-85 of the PISA 2003 Data Analysis Manual.

<sup>20</sup>See Annex A8 of *PISA 2006: Science Competencies for Tomorrow's World, Vol. 1*.

<sup>21</sup>For technical features underlying the estimation procedure of HLM see Woltman *et al.* (2012), Hoofman (1997) and Randenbush and Byrk (2002). Note that due to computational limitation the Balanced Repeated Replication using the 80 population weights provided in the datasets and a Fay's adjustment of 0.5 is not possible to implement under an HLM model.

**Level 2-Model:**

$$\begin{aligned}\beta_{0s} &= \gamma_0 + \gamma_1 S_s + r_{0s} \\ \beta_1 &= \eta_1 \\ \beta_2 &= \eta_2\end{aligned}\tag{8}$$

The estimations using this HLM model correspond to the 4th case presented in the table below.

**Table 4.1**

	5 plausible values reading/mathematics (2003-2012) Weighted sample							
	1		2		3		4	
<b>Student variables</b>	<b>Reading</b>	<b>Maths</b>	<b>Reading</b>	<b>Maths</b>	<b>Reading</b>	<b>Maths</b>	<b>Reading</b>	<b>Maths</b>
Female	22.17*** (15.67)	-25.48*** (-22.29)	22.03*** (16.92)	-26.24*** (-24.71)	21.92*** (16.71)	-26.4*** (-25.09)	21.23*** (17.81)	-26.91*** (-24.27)
Age	-22.83*** (-9.72)	-22.63*** (-9.73)	-22.14*** (-12.29)	-22.04*** (-10.65)	-18.92*** (-11.07)	-17.24*** (-8.94)	-16.63*** (-9.09)	-15.79*** (-8.02)
8th	26.2*** (4.4)	19.51*** (4.4)	24.52*** (5.9)	(3.81) 15.6***	26.82*** (6.69)	19.05*** (4.97)	26.67*** (6.83)	19.81*** (5.55)
9th	74.92*** (18.33)	69.86*** (3.81)	74.37*** (18.83)	67.44*** (19.19)	75.36*** (19.72)	68.92*** (21.28)	73.83*** (20.93)	69.04 (23.67)
10th	144.32*** (33.07)	142.30*** (3.89)	143.78*** (36.01)	139.97*** (38.32)	145.22*** (37.32)	142.09*** (42.47)	140.75*** (41.56)	141.84*** (46.72)
11th	206.6*** (20.34)	218.19*** (21.54)	203.62*** (23.58)	212.57*** (23.7)	206.24*** (23.75)	216.42*** (24.00)	199.36*** (23.18)	214.79*** (24.04)
Academic track	21.04*** (7.35)	18.6*** (6.67)	21.39*** (8.48)	19.42*** (7.89)	22.94*** (8.91)	21.71*** (8.75)	26.78*** (10.18)	23.56*** (4.49)
<b>Family variables</b>								
Mother - High school	5.48*** (2.98)	7.00*** (3.86)	5.14** (5.14)	6.16*** (3.65)	5.41*** (2.96)	6.55** (3.78)	4.34** (2.79)	5.59*** (3.52)
Mother -More than high school	3.16 (1.18)	8.36*** (3.06)	5.72** (2.24)	9.63*** (3.87)	6.88*** (2.71)	11.28*** (4.59)	4.5*** (2.46)	6.14*** (3.63)
Father - High school	5.15*** (2.67)	6.3*** (3.59)	6.35*** (4)	7.30*** (4.68)	5.81*** (3.66)	6.61*** (4.29)	4.19* (1.89)	8.7*** (4.43)
Father -More than high school	6.87*** (2.86)	10.67*** (5.07)	7.40*** (3.65)	11.54*** (6.64)	6.85*** (3.44)	10.89*** (6.31)	6.22*** (2.48)	10.79*** (4.14)
Parentspt1	12.13*** (2.96)	16.19*** (3.54)	14.88*** (4.25)	16.98*** (4.21)	15.39*** (4.47)	17.74*** (4.54)	12.74*** (17.81)	12.03*** (3.69)
Parentspt2	8.8*** (2.71)	13.44*** (3.199)	11.49*** (3.89)	14.29** (3.69)	13.27*** (4.58)	16.96*** (4.49)	11.68*** (4.15)	11.82*** (3.79)
Cultpossinv	46.81*** (7.08)	25.25*** (4.48)	49.04*** (8.25)	26.14*** (4.96)	45.96*** (7.86)	21.6*** (4.13)	47.73*** (8.56)	23.56** (4.49)
Hedresinv	2.3 (0.44)	-0.37 (-0.06)	3.31 (0.66)	-1.25 (-0.24)	13.77*** (2.45)	11.23** (5.18)	13.99** (2.74)	14.09*** (2.72)
Homepos	9.75*** (10.81)	10.89*** (12.14)	8.88*** (11.10)	10.21*** (12.74)	8.92*** (14.3)	9.77*** (12.26)	6.87*** (8.47)	8.61*** (10.96)

<b>School variables</b>								
% of non government financing	0.18*** (3.71)	0.24*** (4.3)	0.17*** (3.93)	0.22*** (4.74)	0.14*** (3.37)	0.18*** (4.09)	0.16*** (3.57)	0.18*** (4.43)
Small town	0.35 (0.085)	-4.02 (-0.94)	0.31 (0.83)	-2.52 (-0.66)	-0.17 (-0.05)	-3.11 (-0.79)	1.53 (0.39)	-3.16 (-0.96)
Town	1.21 (0.27)	-4.26 (-0.89)	0.58 (0.14)	-3.25 (-0.75)	0.64 (0.16)	-3.12 (-0.73)	3.37 (-0.67)	-2.39
City	8.84* (1.88)	4.38 (0.92)	13.11** (2.28)	4.14 (0.92)	8.92** (2.24)	4.6 (0.98)	12.24*** (2.98)	5.93 (1.44)
Large City	12.55** (1.96)	-1.35 (-0.21)	13.11** (2.28)	-1.18 (-0.19)	14.3** (2.46)	0.65 (0.11)	18.24*** (3.31)	2.18 (0.41)
Percentage of girls	0.15 (0.63)	0.00 (-0.024)	0.13 (0.63)	-0.05 (-0.25)	0.32 (1.63)	0.24 (1.27)	0.34* (1.68)	0.19 (1.15)
Schoolsize	0.006* (1.93)	0.0011 (0.39)	0.006** (2.04)	0.002 (0.83)	0.005* (1.94)	0.001 (0.55)	0.008** (2.68)	0.0034 (1.3)
Scmatedu	2.19* (1.81)	2.68** (2.06)	2.7** (2.23)	3.09** (2.52)	2.15* (1.86)	2.29** (2.25)	2.21 (1.88)	2.09** (2.01)
Stratio	-0.21 (-0.92)	-0.26 (-0.98)	-0.18 (-0.85)	-0.27 (-1.24)	0.11 (0.39)	0.16 (0.63)	0.006 (1.02)	0.24 (0.96)
Classize	0.32* (1.65)	0.68*** (3.49)	0.30* (1.69)	0.69*** (3.76)	0.29* (1.67)	0.68*** (3.85)	- -	- -
Proportion of certified teachers	0.20*** (2.99)	0.25*** (3.5)	0.19*** (3.31)	0.25*** (4.03)	0.056 (1.00)	0.041 (0.69)	0.065 (1.082)	0.049 (0.80)
TCSHORT	-1.39 (-0.75)	-2.27 (-1.43)	-1.74 (-1.06)	-2.25 (-1.62)	-0.55 (-0.36)	-0.49 (-0.42)	-0.91 (-0.625)	-1.11 (-0.88)
Constant (38.39)	565.62*** (38.39)	584.29*** (14.11)	559.11*** (18.24)	581.53*** (16.07)	490.27*** (15.95)	479.34*** (13.49)	467.27*** (16.2)	497.78*** (13.79)
Rsqr	0.49	0.48	0.49	0.48	0.49	0.49	-	-
Year Dummies	No	No	No	No	Yes	Yes	Yes	Yes

Statistically significant at \*10%. \*5%. \*\*\*1%. T-ratios in parenthesis.

From the table above it is denoted some consistency across the different specifications presented but some differences when the reading or mathematics score are taken as the explained variable. We decided to include in the frontier estimation those that are significant at least at 10% level in one of the two test scores. For the frontier estimation we consider three different samples which reflect the input analysis performed above. The first two correspond to the completed sample and the imputed sample not controlled for year adjustments and the third one corresponds to the OLS (3) and HLM (4) analysis controlled for year dummies.

Following Bessent and Bessent (1979) it was imposed that the relationship between inputs and outputs must be such that an increase in inputs is translated in an increase in outputs (e.g.  $\beta_i \geq 0$ ).

The inputs measurements must have non-zero elements, and when such cases occurs an infinitesimal value is assumed.<sup>22</sup>

The inputs were classified in two different categories. By one side we had to define those inputs that correspond to the environment where the student lives, and thus are not a choice of the student. In DEA literature these inputs are labelled as non-discretionary,  $z_i$ , and in our case correspond to: gender, age, mother's and father's level of education, parents' nationality, Index of Home possessions (Homepos) and geographic location. All the other inputs,  $x_i$ , the discretionary, correspond to the ones that are indeed included in DEA linear programming problem and that are presented in the table below. Later, in section 5, when we assess the sources of efficiency, the remaining inputs,  $z_i$ , are then included in the analysis.

**Table 4.2**

	(1)	Samples (2)	(3)&(4)
<b>Students' inputs</b>			
Cultposinv	✓	✓	✓
Hedrepossinv			✓
<b>School inputs</b>			
% of non government financing	✓	✓	✓
Percentage of girls			✓
School size	✓	✓	✓
Scmatedu	✓	✓	✓
Class size	✓	✓	✓
Proportion of certified teachers	✓	✓	

In what concerns the inputs, also to note that the variables regarding the grade and track of studies are essentially categorical variables, and they had to be disregarded from the DEA analysis. Thus it was estimated a pooled frontier, where all the students included in the PISA sample are considered. Although in order to incorporate in the analysis these important students' characteristics, then efficient frontiers are derived grouping just those students who are enrolled in the same grade or track of studies. Through this approach we differentiate between a frontier where every student are compared across each other, to a case where students are compared with the ones who are in a similar academic position.

<sup>22</sup>Some of the indexes in the PISA dataset present negative values, since they are constructed in such a way that the OECD average is 0. DEA standard measurements just assume positive values, therefore positive amounts were added to negative values, following the method by Pastor (1994) and Lovell (1995). Alternatively, other specifications on handling DEA with negative data could be used, as in Portela, Thanassoulis and Simpson (2004).

## 4.1 DEA efficiency scores - Results

Three different DEA were run for the samples (1), (2) and (3)&(4) considering as outputs the scores in reading and mathematics and as inputs the ones that result from the regression based choice performed above. Also to stress that the DEA results presented correspond to estimations in which all the students are included independently of the grade or track of studies they are enrolled in. In the following table is shown the quartile distribution on the efficiency scores,  $\theta_0$ , derived:

**Table 4.1.1**

$\theta_0$ (weighted sample)	1st quartile	Median	3rd quartile	Mean
<b>2003</b>	1.16/1.17/1.15	1.27/1.28/1.27	1.41/1.43/1.41	1.31/1.33/1.31
<b>2006</b>	1.17/1.18/1.16	1.29/1.3/1.28	1.46/1.47/1.45	1.35/1.36/1.33
<b>2009</b>	1.18/1.18/1.15	1.29/1.29/1.27	1.44/1.45/1.42	1.33/1.33/1.31
<b>2012</b>	1.17/1.19/1.17	1.28/1.31/1.29	1.44/1.47/1.45	1.33/1.35/1.33

The three values presented for the efficiency score  $\theta_0$  correspond to the DEA estimation according to the samples (1)/(2)/(3)&(4)

Observing the table above it is noticeable how the efficiency levels of the Portuguese students are rather constant across the years, particularly between 2006 and 2012. From all the four distributional points reported to emphasize how this constancy over time is particularly significant in the first quartile and the median, meaning that around the third quartile the efficiency levels seem to be more unstable during the time period considered. These conclusions are robust whether we consider the non-imputed or the imputed database.

**Table 4.1.2**

Mean $\theta_0$ (weighted sample)	Private schools	Public schools
<b>2003</b>	1.27/1.28/1.24	1.31/1.33/1.31
<b>2006</b>	1.3/1.29/1.27	1.36/1.36/1.34
<b>2009</b>	1.24/1.23/1.23	1.34/1.35/1.32
<b>2012</b>	1.24/1.25/1.22	1.34/1.37/1.34

The three values presented for the efficiency score  $\theta_0$  correspond to DEA estimation according to the samples (1)/(2)/(3)&(4)

**Table 4.1.3**

Mean $\theta_0$ (weighted sample)	7th grade	8th grade	9th grade	10th grade	11th grade
<b>2003</b>	1.75/1.77/1.71	1.56/1.58/1.53	1.41/1.43/1.40	1.23/1.24/1.23	1.08/1.11/1.10
<b>2006</b>	1.79/1.81/1.76	1.52/1.53/1.49	1.37/1.37/1.35	1.23/1.24/1.22	1.16/1.16/1.15
<b>2009</b>	1.54/1.55/1.5	1.57/1.58/1.53	1.39/1.39/1.36	1.24/1.24/1.23	1.10/1.08/1.1
<b>2012</b>	1.62/1.65/1.61	1.5/1.56/1.54	1.37/1.4/1.38	1.25/1.26/1.24	1.11/1.14/1.09

The three values presented for the efficiency score  $\theta_0$  correspond to the DEA estimation according to the samples (1)/(2)/(3)&(4)

When desegregating the results it is denoted that the stagnation of the efficiency seems to be happening across the private and public schools. When the levels of efficiency are disaggregated according to the students' grade, we observe that in the grades correspondent to around 83% of the whole sample, 9th and 10th grade, there is a high stability of the efficiency scores across time. The other grades, namely the 7th show a higher level of instability since 2003.<sup>23</sup> Additionally, it was observed that at the individual level, the correlation between the efficiency scores and the average scores in reading and mathematics is negative. Such fact means that higher the score in PISA test, the lower the value of  $\theta_0$ , and then the more efficient is the individual.<sup>24</sup>

## 4.2 Malmquist Index - Comparing PISA across the years.

Caves, Christiansen and Diewert (1981) proposed the concept of Malmquist Index, allowing for efficiency evaluations of the same DMU across different points in time. This measure is particularly appealing for our analysis, since we are measuring efficiency in PISA considering several years. However, contrary to standard Malmquist Index applications, the data used in this work is not panel data. PISA survey intends to be an accurate representation of the student population, and in each time different samples, with different sizes, of students between 15 and 16 years old are considered. This fact makes us use a modified version of the Malmquist Index presented by Camacho and Dyson (2006) [CDMI].<sup>25</sup> In their formulation the authors intended to compare the efficiency performance across different groups. In our specific case, the samples of each of the years in analysis were considered as different groups and Malmquist indexes were computed for 2003-2006, 2006-2009 and 2009-2012. The Malmquist Index has also the interest of allowing to disaggregate between two possible sub-components: 1. Catch up component, standing for the measurement of the movement of the inefficient DMU's to the frontier (efficiency change); 2. Frontiers shifts, namely measuring how from one period to the other the technology allows that with the same inputs higher level of outputs are obtained (technical efficiency).<sup>26</sup>

---

<sup>23</sup>To test for the possibility that the efficiency scores are being driven by top achieving students, the DEA's for cases (2) and (3) were computed, considering just those students whose average score between reading and mathematics is below the 95th percentile for their grade. The results in Appendix A<sub>6</sub> show that the inefficiency levels are lower, but the stability of the results remain as in the tables presented.

<sup>24</sup>A graphical representation of this same relation can be found in Appendix A<sub>7</sub>.

<sup>25</sup>CDMI, Camacho and Dyson Malmquist Index, following the nomenclature by Crespo-Cebada, Chaparro and Santín (2009).

<sup>26</sup>Note that originally, CDMI is used by Camacho and Dyson (2006) under constant returns to scale. Like Portela, Camacho and Keshvari (2012) we use it in variable returns to scale DEA. For a discussion on returns to scale implications on Malmquist Index measurements see Portela and Thanassoulis (2010) and Grifell-Tatj and Lovell (1995).



$$CMI(x_t; y_t; x_{t+1}; y_{t+1}) = \frac{(\prod_{i=1}^{N_t} D^t(x^t, y^t))^{\frac{1}{N_t}}}{(\prod_{j=1}^{N_{t-1}} D^{t-1}(x^{t-1}, y^{t-1}))^{\frac{1}{N_{t-1}}}} \cdot \left[ \frac{(\prod_{i=1}^{N_t} D^{t-1}(x^t, y^t))^{\frac{1}{N_t}}}{(\prod_{i=1}^{N_t} D^t(x^t, y^t))^{\frac{1}{N_t}}} \cdot \frac{(\prod_{j=1}^{N_{t-1}} D^{t-1}(x^{t-1}, y^{t-1}))^{\frac{1}{N_{t-1}}}}{(\prod_{j=1}^{N_{t-1}} D^t(x^{t-1}, y^{t-1}))^{\frac{1}{N_{t-1}}}} \right]^{\frac{1}{2}} \quad (9)$$

$D(x,y)$  stands for the distance function which can be read as the inverse of the efficiency score previous derived:  $\theta_0$ .<sup>27</sup> The term outside the squares brackets represents the efficiency change factor, measuring the efficiency gap between groups (in our particular case different PISA samples in different years)- catch up component. The expressions in brackets, represents the change in the frontier between different PISA samples (years)- frontier shift.<sup>28</sup> Values equal to 1 stand for non progression in the different components of the CDMI, values lower than one stand for a negative evolution and higher than one for a positive one.

As it is observed in (9) the CDMI formulation is composed by ratios of different geometric means, which do not take into account the possibility that we are dealing with a weighted database, as it is the case of the PISA one. This fact leads us propose a modified version of the index named as Weighted Camacho and Dyson Malmquist Index (WCDMI), which differs from the one previously presented since it is composed of weighted geometric means:

$$WCDMI(x_t; y_t; x_{t+1}; y_{t+1}) = \frac{[\prod_{i=1}^{N_t} (D^t(x^t, y^t))^{w_i}]^{\frac{1}{\sum_{i=1}^{N_t} w_i}}}{[\prod_{j=1}^{N_{t-1}} (D^{t-1}(x^{t-1}, y^{t-1}))^{w_j}]^{\frac{1}{\sum_{j=1}^{N_{t-1}} w_j}}} \cdot \left[ \frac{[\prod_{i=1}^{N_t} (D^{t-1}(x^t, y^t))^{w_i}]^{\frac{1}{\sum_{i=1}^{N_t} w_i}}}{[\prod_{i=1}^{N_t} (D^t(x^t, y^t))^{w_i}]^{\frac{1}{\sum_{i=1}^{N_t} w_i}}} \cdot \frac{[\prod_{j=1}^{N_{t-1}} (D^{t-1}(x^{t-1}, y^{t-1}))^{w_j}]^{\frac{1}{\sum_{j=1}^{N_{t-1}} w_j}}}{[\prod_{j=1}^{N_{t-1}} (D^t(x^{t-1}, y^{t-1}))^{w_j}]^{\frac{1}{\sum_{j=1}^{N_{t-1}} w_j}}} \right]^{\frac{1}{2}} \quad (10)$$

<sup>27</sup>In the same way that the higher the value of  $\theta_0$  the lower the level of efficiency, for  $D(x,y)$  the interpretation is reversed, meaning the lower the value of  $D(x,y)$ , the higher the level of inefficiency.  $D(x,y)$  mainly corresponds to the output distance function proposed by Shepard (1970).

<sup>28</sup>The frontier shift component is a geometric mean of the cases when we take the technology in year  $t$  or the technology in year  $t-1$  as the reference. This appears as an advantage compared with standard parametric uses of the Oaxaca-Blinder decomposition which demands the choice *a priori* of a reference year.

From now on, given the coherence in the estimation results between the non-imputed and the imputed samples, just these last ones (2 and 3 4) are considered:

Table 4.2.1<sup>29</sup>

(2)/(3)&(4) (weighted sample)	Malmquist Index	Catch-up	Frontier Shift
<b>2003-2006</b>	$\approx 1.03 / \approx 1.02$	$\approx 0.96 / \approx 0.97$	$\approx 1.04 / \approx 1.06$
<b>2006-2009</b>	$\approx 1 / \approx 1.04$	$\approx 1 / \approx 1$	$\approx 0.99 / \approx 1.04$
<b>2009-2012</b>	$\approx 1.01 / \approx 1.02$	$\approx 0.98 / \approx 0.98$	$\approx 1.03 / \approx 1.03$

The two values presented for the Malmquist Index correspond to the samples (2)/(3)&(4)

The results showed on the Malmquist Index and in its two sub-components intend to infer on the first and second reasons presented in section 2 as possible drivers of the evolution in the Portuguese PISA results, which correspond to shocks in the productivity factor  $A_{ijs}$  (equation 2). In order to assess if this possible technological shock really explains the evolution in PISA scores, the analysis of the Malmquist Index should be put together with the own evolution in PISA scores depicted in table 3.2.

From 2003 to 2006 we observe contradictory movements between the two sub-components that constitute the Malmquist Index. If by one side the value lower than one for the catch-up factor indicates a deterioration in the global level of efficiency (e.g. on average the students are further way from the frontier), by the other side we witness a positive movement of the frontier shift, denoting a higher capacity of the students to transform inputs into outputs. Although to note that the size of this frontier shift effect seems to depend on the sample used, resulting that the total Malmquist Index show a value comprehended between 1.02 and 1.03. Recalling the evolution of the scores between these two same years, we observe a fall in the reading and mathematics results. The results provided by the Malmquist Index analysis do not show a total deterioration of efficiency, then the evolution in the productivity factor  $A_{ij}$  cannot be in the core of the explanation for the evolution of the Portuguese PISA results.

Between 2006 and 2009 we observe that the catch-up factor remains rather constant in both cases (2 and 3&(4)). On the frontier shift side, and depending on the sample used, we observe that we may have a movement in the frontier, leading that the total Malmquist Index is lower bounded by a value of 1 and upper bounded by a value of 1.04. In this period we observe a significant increase in reading and mathematics PISA scores. The Malmquist Index is not able to give a definitive answer for this case, since the evidence on the size of productivity evolution depends on the case considered, and consequently it is not totally clear if it is enough to sustain the evolution in PISA scores. In the last period from 2009 to 2012, we note a slight deterioration in the catch-up component, compensated by a positive evolution in the frontier shift element. When they are aggregated, the Malmquist Index shows a small positive evolution in both cases.

<sup>29</sup>The same table but considering just those students whose average score between reading and mathematics is below the 95th percentile for their school grade is in Appendix A<sub>6</sub>.

Recalling that we observe a small progression in PISA scores in both subjects during these three years, it seems that the evolution in productivity fits the evolution observed in the PISA results.

Across the different Malmquist indexes we note that the frontier shift component has contributed to the improvement of the students' efficiency, although the catch up factor has been having a negative contribution. These two facts put together provide evidence that the progression in efficiency has been driven by the positive evolution of the efficient students rather by the progression of the inefficient ones.

The results observed in the Malmquist Index are not conclusive for the evolution in PISA scores during the PISA waves that comprehend the period between 2003 and 2006 and show to be more insightful between 2006 and 2012. These conclusions make us explore the third possibility presented before, particularly if the evolution in the inputs levels may be a strong justification behind the evolution in the PISA results during this time span.

### 4.3 Inputs Change

To complement the results on the compared analysis between the results of the Malmquist Index and the results in PISA scores, we explore the increase in the inputs as the possible driver for the progression in PISA scores. To test for this possibility we have used what we named an Average Input Quantity Index (AIQI), which captures the change in the weighted mean of each input from one PISA survey to the other:

$$AIQI = \frac{\frac{\sum_{i=1}^{N_t} w_i input_i}{\sum_{i=1}^{N_t} w_i}}{\frac{\sum_{j=1}^{N_{t-1}} w_j input_j}{\sum_{j=1}^{N_{t-1}} w_j}} \quad (11)$$

From all the inputs pointed at the begging of section 4, we have focused in those ones which proved to be strongly significant in all of the three regressions performed. The index is calculated for  $N_t$  students in year t and  $N_{t-1}$  students in year t-1.  $w_i$  and  $w_j$  correspond to the respective final weights attributed to the students by PISA in each year. If the index is higher than 1, we have an increase in the average amount of input from one year to the other. The opposite conclusion can be inferred if this index reveals to be lower than one.

Table 4.3.1<sup>30</sup>

	AIQI		
	2003-2006	2006-2009	2009-2012
Grade	0,97	1,01	1,00
Track of studies <sup>31</sup>	0,97	0,97	0,99
MISCED (0-5)	0.93	1.25	0.93
FISCED (0-5)	0.98	1.14	1.11
Parentspt1	1.33	1.32	0.99
Parentspt2	0.96	0.96	0.99
Cultpossinv	0.96	1.10	0.99
Homepos	0.99	1.05	0.96
Non government financing	0.99	1.04	1.11
Urbanicity (0-5)	0.99	1.02	0.98
School size	0.96	0.98	1.04
Class size	1.00	0.93	1.01
SCMATEDU	0.92	1.08	1.13

In the table above, we can identify the evolution of thirteen inputs from 2003 to 2012. From 2003 to 2006 we can denote that out of the thirteen evaluated just one registered a positive evolution. The opposite movement is seen from 2006 to 2009 when nine of the inputs show values for the AIQI higher than one. Finally from 2009 to 2012 we see a mixed behaviour with five of the thirteen inputs studied increasing, four partially stagnated ( $AIQI \approx 0.99$ ) and the remaining ones decreasing. The evolutions just described must be compared with the evolutions of the PISA scores during the period considered, evaluating in which way the input variations match, or not, the PISA scores evolution. Recalling table 3.1 we observe that from 2003 to 2006 the PISA scores fell, which was followed by a large increase from 2006 to 2009 and to a more modest evolution from 2009 to 2012. The shape of this evolution seems to fit in the pattern of the input progression previously described.

Combining what was presented in sub-section 4.2 and 4.3, it is possible to summarize the conclusions on the reasons behind the evolution in the Portuguese PISA scores: 1. Between 2003 and 2006 despite the evolution in productivity, this has been contradicted by a general fall in the inputs level which lead to the deterioration of the PISA results; 2. Between 2006 and 2009 it may have happened a small technological shock in the education production function which coincided with a general increase in the input levels, which appears as a strong reason to justify the significant progression in PISA scores; 3. From 2009 and 2012, we witness a positive evolution in the frontier shift factor, contradicted, in part, by a regression in the catch-up factor. We observe as well as mix behaviour of inputs, with ones increasing and others decreasing. These patterns show that all these effects put together contributed to the small progression registered in PISA results during this period.

<sup>30</sup>The table presents the results on the imputed sample.

<sup>31</sup>Measured as the percentage of the curriculum that corresponds to the standard academic track, using the courses syllabus provided by the Ministry of Education.

#### 4.4 Enrolment grade and track of studies - Evidence from a separated frontier analysis

In the analysis presented in sub-section 4.1, a unique frontier is constructed for each year, pooling every student in the sample, independently of their academic grade or track of studies. In this subsection we intend to take these two factors in consideration and divide the sample into different sub-groups and run individual frontiers for each sub-group. Thus, we intend to construct frontiers in which each student is compared with the peers that are more alike with them in academic terms, and observe if the conclusions on the evolution of the Portuguese scores taken in the previous section remain. In the sample division according to students' grade we have decided to consider two sub-groups: 1. The students in the lower secondary education, which corresponds to the ones that are enrolled between the 7th and 9th grade; 2. The students who are studying in the upper secondary education, which corresponds to the 10th and 11th grade. Considering the track of studies, two alternative divisions are performed: 1. The students who opted for a vocational track of studies; 2. The students who followed a standard academic school track. As before we initially report the distribution of efficiency scores for the different sub-groups from 2003 to 2012.

Table 4.4.1

$\theta_0$ (2)/(3)	Lower secondary				Upper secondary			
weighted sample	1st Quartile	Median	3rd Quartile	Mean	1st Quartile	Median	3rd Quartile	Mean
2003	1,15/1,14	1,27/1,26	1,42/1,41	1,30/1,29	1,13/1,12	1,22/1,21	1,32/1,31	1,24/1,22
2006	1,19/1,16	1,32/1,30	1,48/1,45	1,36/1,33	1,13/1,11	1,22/1,21	1,31/1,31	1,23/1,22
2009	1,19/1,16	1,32/1,30	1,46/1,42	1,34/1,31	1,14/1,12	1,22/1,21	1,32/1,31	1,24/1,22
2012	1,24/1,23	1,39/1,37	1,56/1,53	1,42/1,40	1,15/1,12	1,24/1,22	1,35/1,32	1,25/1,23

The two values presented for the efficiency score  $\theta_0$  correspond to the DEA estimation according to the samples (2)/(3)&(4)

Table 4.4.2

$\theta_0$ (2)/(3)	Vocational track				Academic track			
weighted sample	1st Quartile	Median	3rd Quartile	Mean	1st Quartile	Median	3rd Quartile	Mean
2003	1,10/1,06	1,21/1,15	1,33/1,28	1,22/1,18	1,15/1,15	1,27/1,26	1,41/1,40	1,31/1,30
2006	1,09/1,07	1,20/1,18	1,34/1,31	1,25/1,23	1,18/1,16	1,30/1,28	1,47/1,45	1,35/1,33
2009	1,09/1,07	1,21/1,19	1,39/1,36	1,26/1,23	1,17/1,15	1,28/1,26	1,43/1,40	1,32/1,30
2012	1,12/1,10	1,23/1,23	1,43/1,41	1,30/1,28	1,18/1,16	1,29/1,247	1,44/1,42	1,33/1,31

The two values presented for the efficiency score  $\theta_0$  correspond to the DEA estimation according to the samples (2)/(3)&(4)

The tables above denote the distribution of the efficiency scores according to the sub-groups previously designed. In the table regarding the students' grade note that the lower secondary students seem to be less constant in their efficiency scores than the ones in upper secondary education. This is particularly significant between 2009 and 2012. When analysing the distribution according to track of studies the conclusions are more complex, with a more inconstant evolution of the efficiency scores in both tracks between 2003 and 2006, but later from 2009 to 2012 the efficiency scores in the vocational track are the ones which are less constant. From 2006 to 2009 both tracks seem rather constant. In this division of the students according to their school track also to notice the surprising result that the students enrolled in the academic track are on average more inefficient when compared than the ones in the vocational track. Although we should

remember that under the academic track we have several students (in the 7th and 8th grade) which already had at least one grade retention, and are probably these students who are driving these results.

As before a more enriching analysis on the evolution of the efficiency scores can be performed. Again following the Weighted Camacho Dyson Malmquist Index (WCDMI), we disentangle the evolution in the efficiency scores in the component that correspond to a catch up movement of inefficient units to the frontier and the component that corresponds to changes in the frontier.

Table 4.4.3

$\theta_0$ (2)/(3)	Lower secondary			Upper secondary		
weighted sample	Malmquist Index	Catch-up	Frontier Shift	Malmquist Index	Catch-up	Frontier Shift
2003-2006	1,07/1,09	0,99/0,98	1,08/1,12	1,01/1,02	1,006/1,005	1,003/1,01
2006-2009	1,03/1,09	0,99/0,94	1,04/1,11	0,97/0,98	0,98/0,98	0,99/1,00
2009-2012	1,03/1,03	0,98/0,98	1,05/1,05	1,00/0,99	0,99/0,99	1,01/0,99

The two values presented for the Malmquist Index correspond to the samples (2)/(3)&(4)

Table 4.4.4

$\theta_0$ (2)/(3)	Vocational track			Academic track		
weighted sample	Malmquist Index	Catch-up	Frontier Shift	Malmquist Index	Catch-up	Frontier Shift
2003-2006	0,97/0,98	0,98/0,97	0,99/1,02	1,00/1,02	0,94/0,96	1,06/1,07
2006-2009	1,05/1,10	0,97/0,97	1,07/1,12	1,00/1,00	1,00/1,00	1,00/1,00
2009-2012	0,96/0,97	0,93/0,93	1,03/1,04	1,02/1,03	1/0,99	1,02/1,03

The two values presented for the Malmquist Index correspond to the samples (2)/(3)&(4)

In the division according to the grade, while in the group concerning the upper secondary education, we observe stable values not just for the Malmquist Index, but also for its two sub-components, by the contrary when we focus on the students in the lower secondary education, the frontier shift element shows strong positive variations across the time, with strong progresses from 2003 to 2009. These movements in the frontier shift seem to be in the core of the changes in the Malmquist Index observed for this same group.

In what regards the division according to the track of studies, from 2003 to 2006 the main movements are observed in the academic track with the progress in the frontier shift being contradicted by the opposite movement in the catch up component. From 2006 to 2009 a constant evolution is observed on the academic track, and a high effect from a frontier shift is denoted in the vocational track. Finally from 2009 to 2012 both tracks witness positive evolution in their frontiers, with this movement being contradicted by a negative evolution in the catch up movement in the vocational track.

Overall we denote that the frontier shift component plays a more significant role in the positive efficiency evolution, meaning the capacity of the students to transform the inputs they are given into higher academic results. This is true for the lower secondary education, with particular incidence between 2003 and 2009, in the academic track between 2003 and 2006 and in the vocational track from 2006 to 2012.

In general, the catch up component presents deterioration values or remains relatively constant for all the different cases studied. This fact shows that it seems not to exist a sustained progress towards a more homogeneous distribution of efficiency (e.g. a higher concentration of the students around the efficient frontier).

These results are in line with what was previously obtained for the case of the pooled frontier, particularly regarding the results presented on table 4.2.1, where it is clear that the positive contribution for the Malmquist Index is given by the frontier shift and not by the catch up component.

## 5 The reasons for students' efficiency

Besides focusing on the reasons that sustain the evolution in the PISA results, the second goal of the present work is to use the DEA efficiency scores to explain why some students are more efficient than others. This is the purpose of subsection 5.1, where we evaluate in which way some environmental factors, such as the gender or the parents' qualification level, can drive the students' ability to generate higher academic outputs from the inputs they are given.

Recognizing the importance of the school inputs in the students' efficiency level we look at two further questions: 1. Identify the students who between 2003-2012 used the school inputs in a more efficient way, pointing out the students who have taken more advantage of the school system to achieve higher academic results; 2. Characterize the students who can be less affected by the reallocation of two particular resources, class size and school size, contributing for the discussion on these two particular school policies.

### 5.1 Environmental factors to explain efficiency - Second stage DEA

To perform the analysis on the reasons behind students' efficiency we come back to the input division presented in section 4, where we divided the inputs in discretionary and non discretionary. Thus in this second-stage DEA, we infer on how the efficiency scores obtained in equation (5) are explained by the environmental factors,  $z_i$ .

$$\theta_i = f(z_i; \mu_i) \quad (12)$$

Instead of following a standard OLS approach or a tobit censored models (since the efficiency scores are lower bounded at one) we follow the methodology proposed by Simar and Wilson (2007) who used a framework for correcting for the expectable correlation between the non-discretionary inputs,  $z_i$ , and the error term  $\mu_i$ . This methodology is performed through two possible algorithm specifications, although in our specific case just the first one is used.<sup>32</sup>

---

<sup>32</sup>A detailed step by step exposition of the algorithm can be seen in Appendix A8. The second algorithm proposed by Simar and Wilson was not applied due to the large size of the samples used.

Table 5.1.1<sup>33</sup>

$\theta_0$ , DEA efficiency scores, (2003-2012) Weighted sample						
	Coef	(2) Lower B.	Upper B.	Coef	(3)&(4) Lower B.	Upper B.
Female	-0,07** (-53.97)	-0.08	-0.06	-0,07*** (-48.89)	-0.08	-0.05
Age	-0.03*** (-13.15)	-0.05	0.00	-0,03*** (-13.99)	-0.05	-0.01
Mother - High school	-0,06*** (-26.82)	-0.07	-0.04	-0,06*** (-27.94)	-0.08	-0.04
Mother - More than high school	-0,05*** (-18.5)	-0.06	-0.03	-0,04*** (-17.91)	-0.06	-0.03
Father - High school	-0,06*** (-27.61)	-0.08	-0.03	-0,06*** (-29.20)	-0.08	-0.04
Father- More than high school	-0,03*** (-13.09)	-0.05	-0.01	-0,03*** (-14.16)	-0.06	-0.01
Parents pt1	-0,15*** (-46.36)	-0.18	-0.12	-0,17*** (-51.51)	-0.2	-0.14
Parents pt2	-0,16*** (-59.90)	-0.19	-0.13	-0,17*** (-64.98)	-0.2	-0.15
Homepos	-0,08*** (-102.70)	-0.09	-0.08	-0,07*** (-90.86)	-0.08	-0.07
Small town	-0,07*** (-27.25)	-0.09	-0.05	-0,05*** (-17.91)	-0.07	-0.02
Town	-0,07*** (-25.93)	-0.09	-0.04	-0,04*** (-16.90)	-0.07	-0.02
City	-0,14*** (-47.72)	-0.17	-0.11	-0,11*** (-39.70)	-0.15	-0.09
Large City	-0,13*** (-31.47)	-0.16	-0.09	-0,11*** (-27.47)	-0.15	-0.07
Constant	2.6*** (69.22)	2.22	2.93	2.53*** (67.14)	2.21	2.87
Sigma	0,3*** (444.89)	0.29	0.3	0,29*** (427.09)	0.28	0.3
Year Dummies	Yes			Yes		

Statistically significant at \*10%, \*5%, \*\*\*1%. T-ratios in parenthesis. The upper and the lower bound values correspond to the 95% confidence interval of the Simar and Wilson 1st algorithm.

From the results presented, we note the high significance that the majority of the non-discretionary inputs seem to have on the efficiency scores. It is observed that girls, from high qualified Portuguese parents who live in wealthier families and in cities tend to have better levels of efficiency<sup>34</sup>. These results help to understand that the students who have a more favourable socio-economic background are the ones who are more able to transform a given set of inputs into higher outputs. If we combine these results with the ones previously obtained in the regressions presented in sub-section 4.1, we can conclude that the factors that explain absolute measures of academic achievement (such as PISA scores) are similar to the reasons that help to explain relative measures of efficiency (such as the efficiency scores,  $\theta_i$ ).

In previous second stage methodologies, such in Afonso and St Aubyn (2005), the results

<sup>33</sup> A robustness check for this regression can be seen in columns 1 and 2 of the table in Appendix 9.

<sup>34</sup> A negative coefficient is read as a positive contribution to efficiency, since lower the value of  $\theta_0$  higher the efficiency level.



from the regressions above are used to correct the initial efficiency scores. We have decided not to follow this strategy, following Cordero, Pedraja and Santín-Gonzalez (2009), which show, by Monte Carlo simulations, that this kind of approaches behave poorly.<sup>35</sup> In alternative we also explored a First Stage DEA including non-discretionary inputs following the model proposed by Ruggiero (1996). The results obtained by this specification and its interpretation can be seen in Appendix A<sub>9</sub>.

## 5.2 Efficiency and school inputs

### 5.2.1 Who takes more advantage of school inputs?

In previous uses of DEA in educational contexts, such as in Kirjavainen and Loikkanen (1998) or Montén and Thater (2010), when deriving the efficiency scores the authors assume several models, with different combinations of inputs and outputs. Following a similar approach in this sub-section we re-run the DEA samples (2) and (3)&(4) but this time just including the school inputs, obtaining once again individual efficiency scores, whose distribution is given as:

**Table 5.2.1.1**

$\theta_{school_0}$ (2)/(3)&(4) (weighted sample)	1st quartile	Median	3th quartile	Mean
(2)	1.20	1.33	1.5	1.38
(3)&(4)	1.2	1.32	1.48	1.37

Distribution of the efficiency scores  $\theta_{school_0}$  considering all the period 2003-2012.

The table above regards data for the 4 samples in analysis (2003-2012). Using these same results, we infer on the reasons that explain these efficiency scores, regressing them on the variables that define the environment where the student lives, represented by the non-discretionary variables,  $z_i$ . Through this analysis we intend to pinpoint who are the students that given the school inputs are more able to achieve higher scores.

$$\theta_{school_i} = f(z_i; \mu_i) \quad (13)$$

As in the previous sub-section this regression is performed using the Simar and Wilson (2007) 1st algorithm, since the efficiency scores are once again truncated at one.

---

<sup>35</sup>Quoting the authors: "Two stage models obtains the worst results due to its own structure which is focused on identifying external variables that really have influence on the results of production rather than worrying about how to construct a boundary frontier to take them into account."

Table 5.2.1.2<sup>36</sup>

$\theta_{school_0}$ , DEA efficiency scores, (2003-2012) Weighted sample						
	Coef	(2) Lower B.	Upper B.	Coef	(3)&(4) Lower B.	Upper B.
Female	-0,08*** (-66.60)	-0.09	-0.07	-0,07*** (-60.99)	-0.08	-0.06
Age	-0.03*** (-14.89)	-0.05	-0.01	-0.03*** (-16.09)	-0.05	-0.01
Mother - High school	-0,05*** (-28.34)	-0.07	-0.04	-0,05*** (-28.65)	-0.07	-0.04
Mother - More than high school	-0,04*** (-17.93)	-0.06	-0.03	-0,04*** (-17.17)	-0.05	-0.03
Father - High school	-0,06*** (-30.92)	-0.08	-0.04	-0,06*** (-32.36)	-0.08	-0.04
Father- More than high school	-0,03*** (-14.81)	-0.05	-0.01	-0,03*** (-14.22)	-0.05	-0.01
Parents pt1	-0,14*** (-45.34)	-0.17	-0.11	-0,16*** (-52.24)	-0.18	-0.13
Parents pt2	-0,14*** (-56.69)	-0.17	-0.12	-0,16*** (-64.55)	-0.18	-0.13
Homepos	-0,11*** (-141.05)	-0.12	-0.10	-0,10*** (-139.09)	-0.11	-0.1
Small town	-0,07*** (-28.63)	-0.09	-0.05	-0,05*** (-20.95)	-0.07	-0.02
Town	-0,07*** (-28.78)	-0.09	-0.05	-0,05*** (-20.73)	-0.07	-0.03
City	-0,14*** (-52.01)	-0.17	-0.12	-0,12*** (-43.76)	-0.14	-0.9
Large City	-0,14*** (-37.54)	-0.18	-0.11	-0,11*** (-30.85)	-0.15	-0.08
Constant	2.87*** (81.63)	2.54	3.18	2.82*** (83.31)	2.52	3.11
Sigma	0,3*** (489.20)	0.28	0.3	0,28*** (494.88)	0.28	0.29
Year Dummies	Yes			Yes		

Statistically significant at \*10%, \*5%, \*\*\*1%. T-ratios in parenthesis. The upper and the lower bound values correspond to the 95% confidence interval of the Simar and Wilson 1st algorithm.

From the table above, the negative coefficients in variable such as Homepos and the parents' level of qualification for example, indicate that individuals whose level of socio-economic condition at home is more favourable have a lower value of  $\theta_{school_0}$  (e.g. a value closer to one, and then are more efficient). Recalling that these levels of efficiency were derived taking into consideration just those inputs related to the school, we can then infer that students whose socio environment is better tend to take more advantage of the school inputs to which they are exposed. This analysis, particularly using a wide time span of 9 years, is relevant to understand who are the students and families who are using the education system of the country in a more positive way, signalling those ones, that by the contrary, are not.

<sup>36</sup>A robustness check for this regression can be seen in columns 2 and 3 of the table in Appendix 9.

### 5.2.2 School inputs allocation: Class size and School size

In this last section we aim to analyse how the efficiency scores previously derived can contribute to the discussion of particular school policy issues. To implement this analysis we extend our DEA model to incorporate the concept of slacks. In section 4, the linear programming problem was solved using a one phase DEA problem. In this unique phase, we approach the concept of radial efficiency, which is reflected by the efficiency score,  $\theta_0$ , whose values higher than 1 suggest that the outputs can be expanded given the inputs available. However this definition does not cover another dimension of efficiency related with inputs excess and outputs shortfalls, normally designated in DEA literature as *slacks*:

$$\begin{aligned} s_i^- &= x_0 - \sum_{j=1}^n x_i \lambda_j \\ s_i^+ &= \sum_{i=1}^n y_i \lambda_i - \theta_0 y_0 \end{aligned} \tag{14}$$

In our case we focus exclusively in the inputs slacks ( $s^-$ ), given by the first equation in (14). These inputs slacks stand for the amount that a certain input could be reduced without hurting the output of the individual, and are normally treated in DEA literature as a second source of inefficiency.<sup>37</sup> We choose to study the slacks related to these two inputs, since they have been particularly important in the discussion of school policy. In the Portuguese case, school size is a relevant question given that during the last years many schools with less of 21 students have been closed, and school administration has been concentrated all over the country.<sup>38</sup> The class size input was chosen given it is for long a topic of discussion in what concerns school policy.<sup>39</sup> To obtain these slack variables, it was necessary to run the linear programming problem, but now a two phase problem was considered, adding the slacks measurement to the previous specification in equation (5):

---

<sup>37</sup>This source is normally labelled as 'non-radial efficiency' or 'mix inefficiency'. We use the slacks as a way to identify the students who have higher slacks in the inputs related to school size and class size. Thus, higher the slack value means that the student maybe enrolled in a larger or smaller class or in a larger or a smaller school and his performance would remain constant.

<sup>38</sup>According to 2010 data from the Portuguese ministry of education, since 2005 3.200 schools were closed, mainly elementary schools in small villages in the country. In 2013, 67 large administrative bodies were created, whose number of students under their responsibility can amount to more than 3.500.

<sup>39</sup>According to OECD 2012 Education at a glance on Portugal between 2000 and 2010 the average class size decreased 34% in the country.

$$\begin{aligned}
& \max \theta_0 + \epsilon \left( \sum_{j=1}^m s_i^- + \sum_{r=1}^s s_i^+ \right) \\
& \text{s.to :} \\
& \theta_0 y_0 \leq \sum_{i=1}^n y_i \lambda_i - s_i^+ \\
& x_0 \geq \sum_{i=1}^n x_i \lambda_i + s_i^- \\
& \sum_{i=1}^n \lambda_i = 1 \\
& \lambda_i \geq 0, s_i^- \geq 0, s_i^+ \geq 0
\end{aligned} \tag{15}$$

This problem was solved for m inputs, s outputs which correspond to the ones previously used on the DEA problem in (5). We collected the slacks associated to the school size and class size inputs, with the purpose of observing how the environmental variables,  $z_i$ , may affect them. The study of the slacks has received less attention in DEA literature, being an exception the work by Fried, Lovell, Schmidt and Yaisawarng (2002). Given that slacks are bounded at zero the first intention would be the use of a Tobit censored model. However, and once again, recalling the work from Simar and Wilson (2007), the authors point out that the approach proposed by them can be applied to cases where slacks are the dependent variable. Again the 1st algorithm by Simar and Wilson was used to perform the following estimation:<sup>40</sup>

$$s_i^- = f(z_i; \mu_{ij}) \tag{16}$$

The results from this estimation using both samples can be observed in the following tables:

**Table 5.2.2.1** <sup>41</sup>

Class size slack, (2003-2012)						
Weighted sample						
		(2)			(3)&(4)	
	Coef	Lower B.	Upper B.	Coef	Lower B.	Upper B.
Female	1.44*** (44.76)	1.20	1.67	1.53*** (45.08)	1.32	1.76
Age	0.003 (0.05)	-0.41	0.39	-0.13*** (-2.28)	-0.50	0.28
Mother - High school	-0.05 (-1.12)	-0.33	0.36	-0.29*** (-6.02)	-0.62	0.05

<sup>40</sup>The detailed explanation of Simar and Wilson 1st Algorithm is in Appendix 7.3, with a note for its use in the slacks case.

<sup>41</sup>A robustness check for this regression can be seen in columns 5 and 6 of the table in Appendix 9.

Mother - More than high school	0.012 (0.22)	-0.26	0.53	-0.057 (-1.04)	-0.39	0.28
Father - High school	0.15*** (3.33)	-0.26	0.53	-0.19*** (-4.08)	-0.57	0.
Father- More than high school	0.25*** (4.66)	-0.15	0.65	0.013 (0.25)	-0.37	0.39
Parents pt1	0.41*** (4.59)	-0.16	1.03	0.64*** (7.35)	0.00	1.31
Parents pt2	0.76*** (9.81)	0.24	1.34	0.72*** (9.47)	0.19	1.31
Homepos	1.06*** (54.50)	0.92	1.19	0.58*** (28.99)	0.46	0.71
Small town	0.53*** (7.01)	0.02	1.03	0.76*** (9.26)	0.28	1.27
Town	0.92*** (12.54)	0.44	1.44	1.64*** (20.47)	1.15	2.16
City	1.46*** (18.45)	0.95	2.01	2.05*** (23.99)	1.50	2.16
Large City	0.22** (2.2)	-0.50	0.90	0.92*** (260.96)	0.22	1.63
Constant	-4.54*** (-5.16)	-10.44	2.2	-2.4*** (-2.63)	-8.99	3.63
Sigma	4.94*** (307.36)	4.81	5.01	4.69*** (260.96)	4.56	4.80
Year Dummies	Yes			Yes		

Statistically significant at \*10%, \*5%, \*\*\*1%. T-ratios in parenthesis. The upper and the lower bound values correspond to the 95% confidence interval of the Simar and Wilson 1st algorithm.

**Table 5.2.2.2** <sup>42</sup>

School size slack, (2003-2012)						
Weighted sample						
	Coef	(2) Lower B.	Upper B.	Coef	(3)&(4) Lower B.	Upper B.
Female	54.97*** (14.4)	24.88	85.32	16.08*** (2.98)	-24.45	55.68
Age	80.29*** (11.97)	23.76	131.20	187.69*** (16.66)	120.17	259.08
Mother - High school	-1.28 (-0.23)	-47.34	39.78	-53.08*** (-6.69)	-113.12	8.28
Mother - More than high school	56.34*** (8.81)	12.16	101.32	74.71*** (8.26)	14.94	133.25
Father - High school	5.27 (0.96)	-49.31	58.00	14.26* (1.84)	-52.59	88.03
Father- More than high school	-24.8*** (-3.83)	-75.77	28.44	-23.57** (-2.58)	-95.32	44.6
Parents pt1	58.83*** (5.38)	-18.53	145.05	204.99*** (13.14)	90.84	331.71
Parents pt2	161.28*** (16.79)	90.11	238.10	373.42*** (27.11)	272.24	490.51
Homepos	88.79*** (38.49)	70.35	105.36	48.933*** (15.17)	25.50	72.10
Small town	115.63*** (10.81)	40.08	196.91	-164.13*** (-11.31)	-252.40	-66.91
Town	506.93*** (48.75)	430.35	584.98	291.69*** (21.16)	203.52	388.35
City	818.08*** (73.59)	738.56	901.09	658.38*** (44.27)	562.09	760.03

<sup>42</sup> A robustness check for this regression can be seen in columns 7 and 8 of the table in Appendix 9.

Large City	531.39*** (41.45)	432.39	632.51	159.89*** (8.98)	18.82	290.02
Constant	-2107.89*** (-19.45)	-2938.25	-1213.76	-4104.86*** (-26.40)	-5198.44	-2957.95
Sigma	626.49*** (301.72)	608.70	643.05	731.69*** (217.44)	705.16	755.83
Year Dummies	Yes			Yes		

Statistically significant at \*10%, \*5%, \*\*\*1%. T-ratios in parenthesis. The upper and the lower bound values correspond to the 95% confidence interval of the Simar and Wilson 1st algorithm.

We note that individuals whose level of Home possessions is higher, live in towns and cities and whose parents are Portuguese have higher levels of slacks, both in what concerns the school size and the class size. This evidence helps us to identify who are the individuals who are able to achieve the same academic outcome being enrolled in a larger or smaller class and school. This indication can be useful when resources are distributed across schools with different types of students, for example when setting the maximum number of students at a given school, where the concentration of students whose level of wealth is high (proxied by the variable, Homepos).

Also to stress the contradictory results regarding the impact of parents' qualifications, depending on the sample used.

## 6 Final Remarks

Every three years when PISA results are released they constitute a source of discussion about how students' performance and schooling systems have been changing over time. This study intended to be a wide assessment and evaluation on the PISA data since 2003. Given the relevance that non-parametric and semi parametric techniques gained in the economics of education literature we based our analysis in the use of these methods, combining them with the particular features of the PISA data. This way, we intended to use the potentialities of the PISA database, contributing for education policy evaluation not just for the Portuguese case, but for the countries who have been participating in the PISA survey.

Two vectors of analysis were followed; proposing a methodology to disentangle the drivers behind the evolution in PISA scores and then assessing why in an education system some students are more able to transform a set of inputs in higher academic results.

Both analysis were performed following an education production function approach combined with the concept of efficiency, identifying the students who can achieve the highest academic outcome possible given the inputs. These students define the efficient frontier which was derived through non-parametric techniques, namely Data Envelopment Analysis (DEA).

On the reasons behind the evolution in PISA scores, it was considered that results could be driven by productivity shocks or by variations in the inputs levels. This last reason revealed to be on the core of the evolution of the Portuguese results, particularly between 2003 and 2009. This result is relevant in the sense that part of the PISA evolution was not due to more or less achieving students but by a progression in the inputs they are exposed, particularly the ones related to the soci-economic conditions at home.

In what concerns progresses in productivity it was considered that it could be driven by an higher capacity of students to transform inputs into outputs (frontier shift factor) or by a more homogeneous distribution of the students around the frontier (catch up factor). The results obtained show that the first component, the frontier shift, was the one which mainly contributed to the positive evolution in the students' productivity.

On the analysis of the reasons that sustain the evolution in PISA scores, a more refined study was performed grouping the students according to their grade and track of studies. In this further step it was observed that the students enrolled in the upper secondary education are the ones whose efficiency scores remained more stable and the students in the lower secondary education denote a consistent and strong evolution. For all the sub-groups formed, once again, the frontier shift explains the evolutions in efficiency observed.

The fact that the frontier shift component presents a systematic positive contribution to the progress in the efficiency and the catch up factor a negative one denote that the progresses in PISA results were mainly driven by the evolution of the most efficient students, meaning that the Portuguese students did not became more homogeneous during the period studied.

We aimed as well to identify the sources of student efficiency and study how it relates to

school inputs. We obtained that the variables which normally explain absolute performance measures (such as PISA scores) are also on the basis students' efficiency. These variables relate mainly with the student's gender and family environment, particularly parents' qualification, place of birth and wealth.

Given the heterogeneity among the students, not just in terms of school results but also in efficiency terms, we questioned how this heterogeneity is reflected in the use of school resources. The results show that students whose socio-economic background is more favourable have been able to use better the school resources to achieve higher academic performance. It is these same students who show to be more indifferent to potential variations in the variables related to class size and school size. This analysis reveals particular insightful in a policy evaluation perspective. By one side it allowed to identify the students who have been using the education system in a more productive way. By the other side it signals that students with a more advantageous socio-economic background maybe enrolled in larger schools and classes, contributing to a better distribution of the school resources and bringing a better balance between costs and benefits in the education system.



## 7 References

- Aderson P., Peterson, N. (1993). "A Procedure for Ranking Efficient units in Data Envelopment Analysis", *Management Science*, 39(10), 1261-1264.
- Afonso, A, St. Aubyn, M. (2005). "Cross country Efficiency of Secondary Education. A semi-parametric Analysis with Non-discretionary inputs", ECB Working Paper.
- Agasisisti, T., Cordero-Ferrera, Jose.M. (2013). "Educational disparities across regions: A multilevel analysis for Italy and Spain", *Journal of Policy Modeling*, 36(6), 1079-1102.
- Bessent A, Bessent W. (1979). "Determining the comparative efficiency of schools through data envelopment analysis", Center of Cybernetic Studies.
- Bowers A, Urick, A. (2011). "Does High School Facility Quality affect Student Achievement?: A Two-Level Hierarchical Linear Model", *Journal of Education Finance*, 37 (1), 72-94.
- Camacho, A., Dyson, R. (2006). "Data Envelopment Analysis and Malmquist indices for measuring group performance", *Journal of Productivity Analysis*, 26(1), 35-49.
- Charnes, A. Cooper, W., Rhodes,E. (1978). "Measuring the efficiency of decision Making units", *European Journal of Operational Research*, 2(6), 429-444.
- Cooper, W., Seinfeld, L., Tone, K. (2007). "Data Envelopment Analysis: A comprehensive Text with Models, Applications, References and DEA-Solver Software", Springer, 2nd Edition.
- Cebada-Crespo, Eva, F., Santín-Gonazález, D. (2009). "Comparing school ownership performance using a pseudo-panel database: a Malmquist-type index approach" 2009.
- Cordero-Ferrera, J., Pedraja-Chaparro, F., Santín-Gonazaález, D. (2006). "Alternative Approaches to Include Exogenous Variables in DEA Measures: A comparison using Monte Carlo", *Computers Operations Research*, 36(10), 2699-2706.
- Deutsch, J., Dumas, A., Silber, J. (2013). "Estimating an educational production function for five countries of Latin America on the basis of PISA data", *Economics of Education Review*, 36, 245-262.

Före, R., Grosskopf, S., Zhang, Z. (1994). "Productivity Growth, Technical Progress, and Efficiency Change in Industrialized Countries", *The American Economic Review*, 85(5), 1033-10.

Farrell, M. (1957). "The measurement of productivity efficiency", *Journal of Royal Statistics Society*, 120(3), 253-290.

Fried, H., Lovell, C., Schmidet S., Yasaiawarng S. (2002). "Accounting for environmental Effects and Statistical Noise in Data Envelopment analysis", *Journal of Productivity Analysis*, 17(1-2), 157-174.

Fried, H.O., Lovell, C.A.K, Schmidt S.S., Lovell C.A. (2008). "The Measurement of Productive Efficiency and Productivity Growth", Oxford University Press.

Grifell-Tatj, E., lovell, C.A.K. (1995). "A note on the Malmquist productivity index", *Economics Letters*, 47, 169-175.

Hanushek, E.(1986). "The economics of schooling: production and efficiency in public schools", *Journal of Economics Literature*, 49(3), 1141-1177.

Hanushek, E. (2003). "The failure of input-based schooling policies", *Economic Journal*, 113(2), 64-98.

Hofmann, D. (1997). "An overview of the Logic and Rationale of Hierarchical Linear Models", *Journal of Management*, 23(6), 723-744.

Johnson, A.,Ruggiero, J (2011). "Non parametric Measurement of Productivity and Efficiency in Education", *Annals of Operations Research*, 221 (1), 197-210.

Karjavainen, T., Loikkanen, H. (1998). "Efficiency Differences of Finnish Senior Secondary Schools: An application o DEA and Tobit Analysis", *Economics of Education Review*, 17(4), 377-394.

Lovell, C., Pastor, J. (1995) "Units invariant and translation invariant DEA models", *Operations Research Letters*, 18(3), 147-151.

Mancebón-Torrubia, M., Calero, J., Choi, A., Ximénez, D. (2010). "Efficiency of public and publicly-subsidized high schools in Spain, Evidence from PISA 2006", MPRA Paper.

Montén. A., Thater, C. (2010). "Determinants of child care provision", IFO working paper.

Nataraja, N, Johnson, A. (2011). "Guidelines for Using Variable Selection Techniques in Data Envelopment Analysis", *European Journal of Operational Research*, 215(3), 662-669.

OECD (2003, 2006, 2009 and 2012), "PISA Data Analysis Manuel", Paris: Organization for Economic Co-operation and Development.

OECD (2003, 2006, 2009 and 2012), "PISA Technical Reports", Paris: Organization for Economic Co-operation and Development.

OECD (2003, 2006, 2009 and 2012), "PISA 2006: Science Competencies for Tomorrows World, Vol. 1.", Paris: Organization for Economic Co-operation and Development.

OECD (2012), "Education at a glance", Paris: Organization for Economic Co-operation and Development.

Pastor, J. (1994). "How to discount environmental effects in DEA: an application to bank branches" Working paper N. 011/94, Depto. De Estadística e Investigación Operativa, Universidad de Alicante, Spain.

Perelman, S., Santín, D. (2011) "Measuring educational efficiency at the student level with parametric stochastic distance functions: an application to Spanish PISA results", *Education Economics*, 19(1), 29-49.

Pereira, M. (2011). "An analysis of the Portuguese Students' Performance in the OECD Programme for International Student Assessment (PISA)", *Economic Bulletin, Banco de Portugal*, Autumn 2011.

Pereira, M., Moreira, S. (2007). "Efficiency of Secondary Schools in Portugal: A Stochastic Frontier Analysis", *Economic Bulletin, Banco de Portugal*, Spring 2007.

Portela, M., Thanassoulis, E., Simpson G. (2004). "Negative Data in DEA: A directional Distance Approach Applied to Bank branches", *Journal of the Operational Research Society*, 55(10), 1111-1121.

Portela, M., Thanassoulis, E. (2010). "Malmquist-type indices in the presence of negative data: An application to bank branches", *Journal of Banking and Finance*, 34(7), 1472-1483.

Portela, M., Camacho, A., Keshvari, A. (2012) "Assessing the evolution of school performance and value added: trends over four years", *Journal of Productivity Analysis*, 39(1), 1-14.

Raudenbush, S., Bryk, A.S. (2002). "Hierarchical linear models: Applications and data analysis", Second edition, Newbury Park, CA:Sage.

Ryan, C. (2013). "What is behind the decline in student achievement in Australia?", *Economics of Education Review*, 37: 226-239

Sena, V. (2003). "The Frontier Approach to the Measurement of Productivity and Technical Efficiency", mimeo

Simar, L., Wilson, P. (2007). "Estimation and inference in two-stage, semi-parametric models of production process", *Journal of Econometrics*, 136(1), 31-64.

Todd, P., Kenneth, W. (2003). "On the specification and estimation of the production function for cognitive achievement", *The Economics Journal*, 113(485), 3-33.

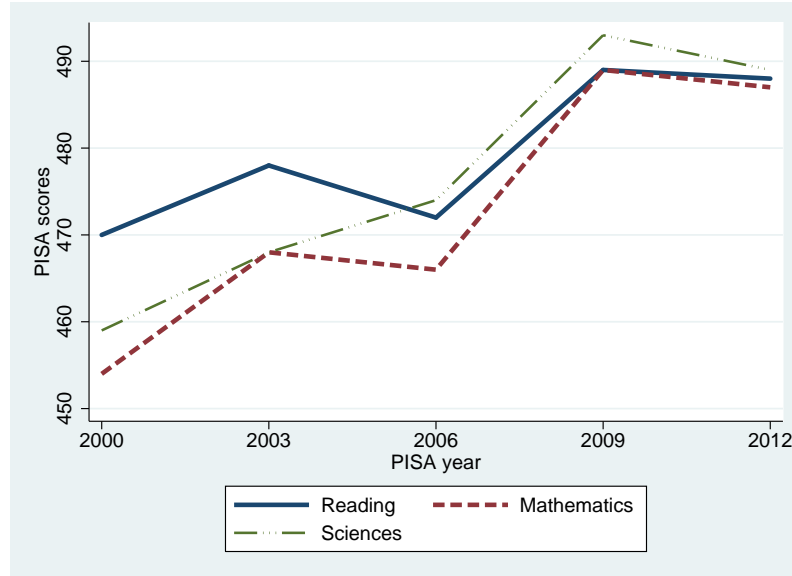
Waldo, S. (2005). "On the use of student data in efficiency analysis - technical efficiency in Swedish upper secondary school", *Economics of Education Review*, 26(2), 173-185.

Woltman, H., Feldstain, A, MacKay,J., Rocchi, M. (2012). "An introduction to hierarchical linear modelling", *Tutorials in Quantitative Methods for Psychology*, 8 (1), 52-69.

Worthington, A. (2001). "An empirical Survey of Frontier Efficiency Measurement Techniques in Education", *Education Economics*, 9 (3), 245-268.

## A Appendix

### A<sub>1</sub> Portuguese average PISA scores 2003-2012

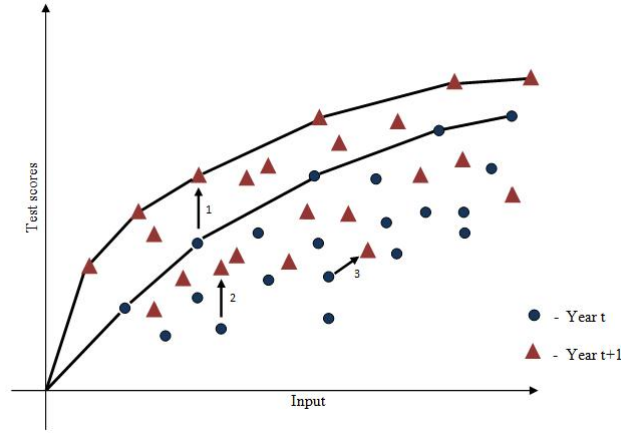


### A<sub>2</sub> General structure of the Portuguese system

	Lower Secondary education			Upper Secondary education	
	7th grade	8th grade	9th grade	10th grade	11 th grade
Standard academic track	✓	✓	✓	✓	✓
CEF - Vocational oriented courses (Ungraded)					
Professional Courses				✓	✓
Technological Courses				✓	✓

The rows correspond to the different tracks and the columns to the different grades covered in the PISA sample. The ticks signals the grades which offer the respective tracks of studies.

### A<sub>3</sub> Three possible explanations to explain the evolution in PISA scores



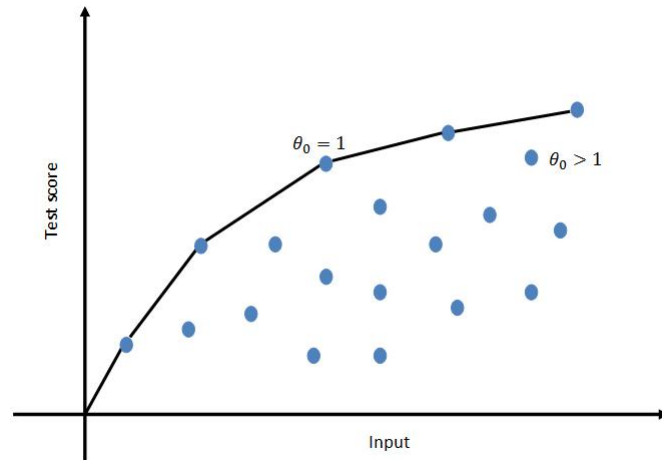
### A<sub>4</sub> Expected maximization algorithm (EM) for data imputation

This formulation follows closely the EM steps presented in *Microeconometrics, Methods and applications* by Cameron and Triverdi (8th edition, chapter 27.5.1, pages 930-932). These steps are applied to a sample composed by  $N$  observations, and from these ones for each variable,  $N_1$  are observed and  $N_2$  are missing:

1. Estimate  $\hat{\beta}$  using the  $N_1$ .
2. Generate  $\widehat{y}_{mis} = X_2\hat{\beta}_2$ .
3. Generate the adjusted values of  $\widehat{y}_{mis}^a = (\widehat{V}^{-1/2}\widehat{y}_{mis}) \odot \mu_m$  of  $\widehat{y}_{mis}$ , where  $\widehat{V}$  is the variance of  $\widehat{y}_{mis}$ ,  $\mu_m$  is a Monte Carlo draw from the  $\mathcal{N}[0, s^2]$  distribution and  $\odot$  stands as element-by-element multiplication.
4. Using the augmented sample obtain a revised estimate of  $\hat{\beta}$ .
5. The steps 1-4 are then repeated for 200 interactions which is sufficient to achieve convergence.

Note: School database was initially imputed and just then using this information it was imputed the student database. In the imputation routines we included the variables, number of books at home, Respres (Index of school autonomy in management of school resources) and Respcurr (Index of school autonomy in curriculum), despite they are not present later in the final estimations performed.

## A<sub>5</sub> DEA graphical representation



## A<sub>6</sub> DEA scores at 95th percentile

As for the standard case when all the students are considered, in this frontier measurement we present the distribution of the efficiency scores as well the results on the malmquist index analysis. These results should be contrasted with the ones presented in tables 4.1.1 and 4.2.1 respectively:

Table A 6.1

$\theta_0$ (2)/(3)&4 (weighted sample)	1st quartile	Median	3rd quartile	Mean
<b>2003</b>	1.16/1.13	1.26/1.24	1.41/1.38	1.32/1.28
<b>2006</b>	1.13/1.12	1.24/1.23	1.41/1.39	1.3/1.28
<b>2009</b>	1.13/1.12	1.24/1.23	1.38/1.37	1.28/1.26
<b>2012</b>	1.14/1.12	1.24/1.23	1.39/1.37	1.29/1.27

The three values presented for the efficiency score  $\theta_0$  correspond to the DEA estimation according to the samples (2)/(3)&(4)

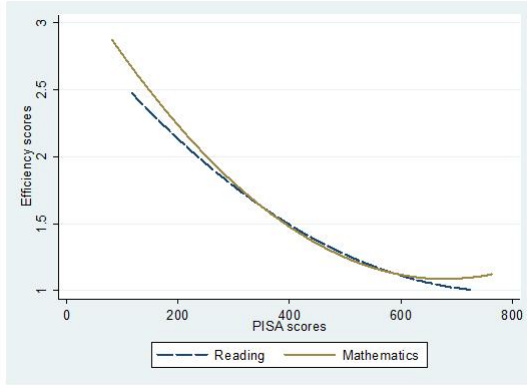
Table A 6.1

(2)/(3)&4 (weighted sample)	Malmquist Index	Catch-up	Frontier Shift
<b>2003-2006</b>	$\approx 0.99 / \approx 1.02$	$\approx 0.98 / \approx 0.97$	$\approx 1.02 / \approx 1.05$
<b>2006-2009</b>	$\approx 0.99 / \approx 1.04$	$\approx 1 / \approx 1$	$\approx 0.99 / \approx 1.03$
<b>2009-2012</b>	$\approx 1.01 / \approx 1.02$	$\approx 0.99 / \approx 0.99$	$\approx 1.02 / \approx 1.02$

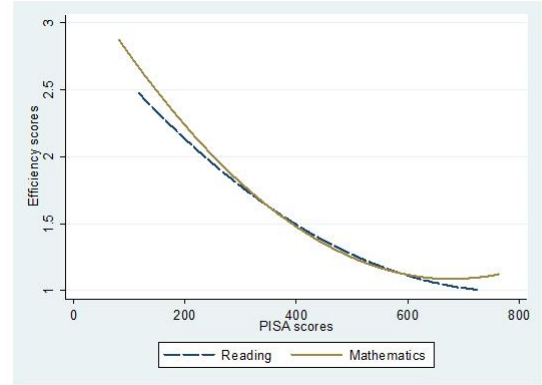
The two values presented for the Malmquist Index correspond to the samples (2)/(3)&4

From the tables above we note that the results previously derived are mainly confirmed by this specific case. This coherence in the conclusions is observed both in the efficiency scores, where the constancy across time is observed, as well in the malmquist indexes, where the results for its different components fit in the conclusions previously derived.

#### A<sub>7</sub> Relation between PISA scores and efficiency scores



Quadratic fitting between efficiency scores and PISA scores in mathematics and reading, sample (2)



Quadratic fitting between efficiency scores and PISA scores in mathematics and reading, sample (3) & (4)

#### A<sub>8</sub> Simar and Wilson Bootstrap algorithm

- 1) Estimation of  $\hat{\theta}_j$  for all DMU's considering the original data.
- 2) By Maximum Likelihood obtain  $\hat{\beta}$  and  $\hat{\sigma}_\epsilon$  from the truncated regression model,  $\hat{\theta}_i = f(z_i; \mu_i)$ .
- 3) Loop over the next three steps for L bootstraps estimates for  $\beta$  and  $\sigma_\mu$ :
  - 3.1 For each j DMU's draw  $\mu_j$  from  $\mathcal{N}[0, \hat{\sigma}_\mu^2]$  distributed with left truncation at  $1 - z_j \hat{\beta}$ .
  - 3.2 Compute  $\theta_j^* = z_j \hat{\beta} + \mu_j$ .
  - 3.3 Come back to the maximum likelihood truncated regression to estimate.  $\theta_j^* = f(z_i; \mu_i)$ .
- 4) Use the bootstrap estimates and the original  $\hat{\beta}$  and  $\hat{\sigma}_\mu$  to construct confidence intervals for  $\beta$  and  $\sigma_\mu$ .

Note: When using slacks as the depend variable, instead of, 1, the left truncation point is 0.



#### A<sub>9</sub> Including non-discretionary inputs in a one-stage DEA model

In 1996 in the work *On the Measurement of efficiency in public sector* John Ruggiero proposed the following modified DEA model to account for environmental factors:

$$\begin{aligned}
& \max \theta_0 \\
& s.to : \\
& \theta_0 y_0 \leq \sum_{i=1}^n y_i \lambda_i \\
& x_0 \geq \sum_{i=1}^n x_i \lambda_i \\
& \sum_{i=1}^n \lambda_i = 1 \\
& \lambda_i = 0 \text{ if } z_i > z_0, \text{ otherwise } \lambda_i \geq 0
\end{aligned} \tag{17}$$

This model proposed by Ruggiero basically restricts for each DMU its reference set according to the non-discretionary variable,  $z_i$ , considered. In our case we considered as the reference non-discretionary variable the value of Home Possession at home. This choice is justified, since this is the variable used to proxy for the students' family wealth, and in the second stage DEA presented in sub-section 5.1 it reveals to be significant to explain the efficiency scores. The results from this type of DEA are presented in the table below:

**Table A 9.1**

$\theta_0$ (2)/(3) (weighted sample)	1st quartile	Median	3rd q uartile	Mean
<b>2003</b>	1.09/1.08	1.19/1.19	1.33/1.32	1.23/1.23
<b>2006</b>	1.08/1.06	1.20/1.18	1.35/1.32	1.25/1.22
<b>2009</b>	1.09/1.07	1.20/1.18	1.34/1.31	1.24/1.21
<b>2012</b>	1.10/1.07	1.21/1.19	1.35/1.33	1.25/1.23

The three values presented for the efficiency score  $\theta_0$  correspond to the DEA estimation according to the samples (2)/(3)&(4)

It is observed that the constancy found in standard DEA models is reproduced in this same case.

# A<sub>10</sub> Robustness table

In the following table are showed the coefficients of the truncated regressions performed but this time including grade and track of studies as controls.

Table A10.1

Robustness table (2003-2012) (Weighted sample)								
	(1) Coef	(2) Coef	(3) Coef	(4) Coef	(5) Coef	(6) Coef	(7) Coef	(8) Coef
8th	-0,11***	-0,11***	-0,11***	-0,11***	1,3***	0,06	277,40***	263,37***
9th	-0,28***	-0,27***	-0,29***	-0,29***	1,59***	0,75***	259,62***	219,44***
10th	-0,51***	-0,49***	-0,51***	-0,5***	2,73***	1,60***	698,34***	829,32***
11th	-0,79***	-0,81***	-0,77***	-0,75***	2,5***	1,17***	826,06***	1031,23***
Academic track	-0,02***	-0,01***	-0,03***	2,35***	2,64***	2,16***	-147,36***	-30,48***
Female	-0,01***	-0,01***	-0,01***	-0,01***	1,25***	1,37***	14,02***	-30,48***
Age	0,07***	0,06***	0,06***	0,06***	-0,35***	-0,38***	-54,61***	-20,10**
Mother - High school	-0,02***	-0,03***	-0,02***	-0,02***	-0,11***	-0,36***	-16,63***	-72,48***
Mother - More than high school	-0,02***	-0,02***	-0,09***	-0,02***	0,00	-0,07	35,66***	40,61***
Father - High school	-0,02***	-0,02***	-0,02***	-0,02***	0,027	-0,28***	-9,41**	-21,73***
Father- More than high school	-0,01***	-0,02***	-0,02***	-0,02***	0,2***	-0,02	-16,56***	-21,3***
Parents pt1	-0,06***	-0,08***	-0,05***	-0,07***	0,28**	0,48***	-11,04	56,82***
Parents pt2	-0,05***	-0,07***	-0,04***	-0,06***	0,58***	0,53***	70,73***	195,24***
HOMEPOS	-0,01***	-0,01***	-0,04***	-0,03***	0,8***	0,36***	40,69***	-1,76
Small town	0,00	0,02***	-0,02	0,01***	0,32**	0,53***	12,24	-243,41***
Town	0,04***	0,05***	0,03***	0,05***	0,64***	1,36***	347,42***	112,73***
City	-0,02***	-0,01***	-0,03***	-0,01***	1,06***	1,69***	626,33***	443,96***
Large City	-0,03***	-0,02***	-0,05***	-0,02***	-0,16	0,55***	384,94***	17,95
Constant	0,79***	0,78***	1,07***	1,14***	-0,93	0,38	214,21**	-519,09***
Sigma	0,23***	0,23***	0,22***	0,21***	4,89***	4,65***	596,02***	682,61***
Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Statistically significant at \*10%, \*5%, \*\*\*1%.